Management plane for differential privacy preservation through smart contracts

Nida Khan, Abdelkader Lahmadi, Zsofia Kräussl, Radu State

Management plane for differential privacy preservation through smart contracts

Nida Khan *, Abdelkader Lahmadi †, Zsofia Kräussl ‡ and Radu State *
*Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg
{nida.khan, radu.state}@uni.lu
†Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
abdelkader.lahmadi@loria.fr
‡Department of Finance, University of Luxembourg, Luxembourg
zsofia.kraussl@uni.lu

***Abstract*—Blockchain has emerged as a novel solution addressing a plethora of industrial issues in domains spanning from financial to educational. However, several challenges restrict the widespread adoption of the technology and data privacy, with throughput and scalability issues, ranks amongst the foremost. In this paper, we introduce a novel privacy management plane which integrates differential privacy to query existing relational databases through the blockchain as well as spearheads the use of blockchain for local differential privacy. The distinguishing feature in the latter is that the privacy management plane gives the data owners the right to perturb their data with the desired privacy budget, while in the former it gives the right to the data curator to change the privacy budget dynamically while answering queries through the blockchain. The paper also includes experimental evaluation of the developed privacy management plane and integrates management operations in it through another smart contract. The paper addresses the issue of GDPR and it's implications in the context of blockchain data, while highlighting the compliance of the proposed implementation.**

***Index Terms*—differential privacy, Laplace, blockchain, Ethereum, smart contract, GDPR**

I. Introduction

The advent of blockchain in 2008 with Bitcoin heralded a novel era of technological innovation pervading primarily the finance domain. The emerging technology expanded to other domains with the passage of time, like the education, healthcare and the supply chain industry among others. However, the envisaged utility of the technology was challenged by inherent bottlenecks like scalability, throughput, data privacy and security. The decentralized distributed database when compared to relational databases scores a lower score in not just performance but ease of use, latency and lack of data deletion and updating [1]. The core features that blockchain technology is leveraged upon are a trustless environment, immutability and transparency, which come at the cost of lack of data privacy, among the other listed challenges. The distributed network with no single entity holding an obligation for the entire network has created the need for new regulations with regards to legality of smart contracts, dispute resolution for transactions taking place through the network and even economic uncertainty regarding the status of cryptocurrency as being analogous to fiat money or a digital token. The technology research firm Gartner has called *blockchain privacy poisoning* [2], which implies insertion of personal data in the public blockchain, as one of the biggest risks facing the organizations as that makes blockchain non-compliant under the European General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA). The requirements by both the laws is that the user data be deleted on need in adherence to the *"the right to be forgotten"*. The immutable feature of the distributed ledger makes it vulnerable to cyber-attacks [3] and there exists a need to conceive a model for utilization of blockchain by organizations whereby the design of the decentralized applications on the blockchain makes it resilient to such attacks.

In the wake of GDPR, CCPA and the vulnerability of privacy-preserving blockchain platforms there is a need to address the issue of privacy using a formal, mathematical model for data privacy, namely differential privacy. There is a necessity to develop a design mechanism for blockchain usage, where the benefits of the technology do not compromise on data privacy. Dash is a privacy-preserving blockchain and Koldner *et al.* proved that it is susceptible to the cluster intersection attack [4]. Monero is another privacy-preserving blockchain that utilizes one-time ring signature scheme, attempting to provide both untraceability and unlinkability, but is subject to temporal analysis [5]. Zcash provides private transactions secured by zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs) but were detected as suffering from an inflation bug [6].

In this paper we focus on the design, development and testing of a novel privacy management plane for preserving the privacy of data using blockchain. We use differential privacy as a privacy-preservation mechanism, which is a formal, mathematical definition of privacy ensuring the preservation of privacy during analyses. We utilize Laplace mechanism to perturb the data and additionally we integrate the privacy management plane with a managing smart contract to facilitate role-based data access in the smart contract of the privacy management plane. We discuss our implementation with respect to GDPR 2016/679, which is concerned with privacy of data in the European Union and the European Economic Area as well as exchange of personal data outside the addressed region.

The rest of the paper is structured as follows: State of the art is given in Section II and the relevant background on differential privacy is given in Section III. The design overview of the privacy management plane is given in Section IV, which gives the functional architecture of the privacy management plane. The implementation of the privacy management plane is discussed in Section V, while the experimental evaluation of the developed privacy management plane is highlighted in Section VI. GDPR and its implications are discussed in Section VII, while the conclusion is given in Section VIII.

II. Related Work

Zyskind et al. construct a data management platform which involves using blockchain to restrict access to data in an off-blockchain storage, with the key to the data stored on the blockchain [7]. Chen et al. propose a decentralized machine learning system, *LearningChain*, which ensures differential privacy [8]. Hassan et al. discuss the privacy issues raised by integration of blockchain with IoT and analyze privacy preservation strategies, including differential privacy, in blockchain-based IoT systems [9]. Yang et al. propose a blockchain-based anonymization process for the data owners while providing the functionality to validate the privacy budget and adapt it to the privacy requirements of the data owners [10]. Dagher et al. proposed *Ancile* for access control and interoperability of Electronic Health Records. They use permissioned blockchain to store the hashes of records and highlighted differential privacy as future work [11]. Khan and Nassar analyze recently proposed privacy-preserving blockchains and emphasized on the need to implement differential privacy to ensure data privacy in blockchains [12]. The present work differs in proposing a novel privacy management plane, which utilizes smart contracts to store the results of queries on the blockchain using differential privacy. Our work also uses smart contracts to implement local differential privacy giving an individual user, the freedom to determine his privacy budget. In accordance with GDPR, the privacy management plane when used in a permissioned blockchain network, can prevent read access to a data record.

III. Background: Differential Privacy

Differential privacy is a formal mathematical definition of privacy, which ensures utilizing data for analysis while preserving privacy [13]–[15]. In the present age where data has become a utility, and extensive analysis of medical records and even online activities is being conducted for research purposes, the significance of individual privacy has increased manifold. The access to data while ensuring a strong privacy mechanism to retain both accuracy of data shared within the limits of privacy is crucial [16]. The objective of differential privacy is to ensure that the analyst remains as unaware about any individual in the target dataset post the analysis as he was prior to the analysis. Thus differential privacy ensures that there is no leakage of data at the individual level in the database. Privacy can be expressed according to the following approaches:

1) **Global privacy.** The need for global privacy arises when an organization releases the database of several people or answers queries on the database, comprising of n number of rows. A *query*, q, from the query space $\mathcal{Q}$ is a function to be applied to the database. The privacy mechanism, $\mathcal{M}$ is an algorithm that can be expressed as [17]:
$$\mathcal{M} : \mathcal{X}^n \times \mathcal{Q} \longrightarrow \mathcal{Y} \tag{1}$$
In equation 1, $\mathcal{X}$ is the *data universe* comprising of the rows of data types, where dataset $x \in \mathcal{X}^n$. $\mathcal{Y}$ is the *output space* of $\mathcal{M}$. The privacy mechanism thus, takes in as input a dataset and a set of queries producing an output string, which is expected to provide answers to queries preserving differential privacy. This approach is henceforth referred to as *global differential privacy* (GDP) in the paper.The privacy mechanism *M* is said to be *ϵ-differentially private* if it satisfies the following for every pair of neighbouring datasets x and x' and every query $q \in \mathcal{Q}$, where $x, x' \in \mathcal{X}^n$ [17]:
$$\forall y \in \mathcal{Y}, \mathcal{P}r[\mathcal{M}(x, q) = y] \leq e^{\epsilon} \cdot \mathcal{P}r[\mathcal{M}(x', q) = y] \tag{2}$$
In equation 2, x and x' differ by only one row.
2) **Local privacy.** The need for local privacy arises when a user discloses his personal information voluntarily. This disclosure assumes that the data owners do not trust the data curator and add noise to their data locally. The privacy algorithm can be expressed as [18]:
$$\forall y \in Range(\pi) : \mathcal{P}r[\pi(v) = y] \leq e^{\epsilon} \mathcal{P}r[\pi(v') = y] \tag{3}$$
It is assumed that the user has a private value v_i in some domain D, then the algorithm π is used to add noise to v_i and is sent as $\pi(v_i)$ to the data curator to derive statistical information from it. Equation 3 depicts the set of all possible output values in $Range(\pi)$ for any input v and v' for the algorithm π, where $\epsilon \geq 0$. This algorithm is formally taken as satisfying *ϵ-local differential privacy* [18]. This approach is henceforth referred to as *Local Differential Privacy* (LDP) in the paper.

Privacy budget is indicated by ϵ and it is used to control the output of the algorithms used in a privacy mechanism. The privacy budget determines how private the output of the algorithm is. Smaller values of ϵ indicate more private data with a loss of accuracy. The value of ϵ is generally taken to be 0.01, 0.1, 0.5 and 0.8 [19].

Sensitivity determines how much noise is to be added to the results in differential privacy. It depends on how much the output can change on the insertion or deletion of a single row in a dataset.

A. Laplace Mechanism

The primary mechanisms to add noise in differential privacy are the Laplace mechanism and the exponential mechanism. The parameter noise is related to the privacy budget and the sensitivity. The Laplace mechanism adds perturbation to the data with noise according to the Laplace distribution in

a numerical output whereas the exponential mechanism is mainly used when the outputs are non-numerical. In this work, we focused on Laplace mechanism since the dataset we employed for evaluation is mainly numerical. The scale of the noise in the Laplace mechanism is dependent on the sensitivity function, divided by ϵ [20].

1) Laplace Distribution.: The Laplace distribution, $Lap(b)$ with a scale b (centered at 0) is the distribution with the probability density function [21]:

$$Lap(x|b) = \frac{1}{2b} exp\left(-\frac{|x|}{b}\right) \tag{4}$$

The variance of the distribution is $\sigma^2 = 2b^2$. In equation 4, $x \in \mathbb{N}^{|\mathcal{X}|}$, where the domain $\mathbb{N}$ denotes the set of all non-negative integers including zero. The most fundamental types of database queries are numeric queries, functions:

$$f : \mathbb{N}^{|\mathcal{X}|} \longrightarrow \mathbb{R}^k \tag{5}$$

Equation 5 depicts queries that map databases to k real numbers. The Laplace Mechanism for any function f given in equation 5 can be stated as:

$$\mathcal{M}_L(x, f(.), \epsilon) = f((x) + (Y_1, ...Y_k) \tag{6}$$

In equation 6, Y_i are random variables drawn from $Lap(\Delta f/\epsilon)$ [21].

IV. Design Overview: Privacy Management Plane

We developed a privacy management plane to assist in the preservation of both local and global differential privacy. The privacy management plane comprises of a smart contract deployed on Ethereum and a web application. The smart contract is deployed on the Ethereum blockchain network through a decentralized application. The smart contract has functions that cater to both global privacy for organizations owning private databases as well as local privacy for individual data owners. The functional architecture of the privacy management plane is given in Fig. 1.

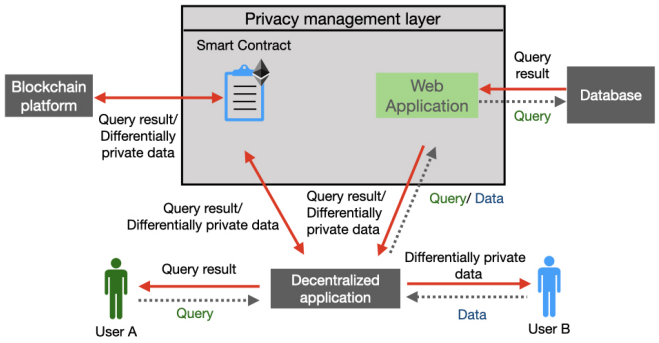

Fig. 1: Functional Architecture of the Privacy Management Plane

The mechanism to achieve differential privacy using the privacy management plane is enumerated below:

- Local Differential Privacy: The user inputs the raw data from the frontend of the decentralized app, which is sent by a POST request to a web application, which adds Laplace noise to the raw data depending on the value of ϵ chosen by the user. The raw data is never stored in the web application. Once the Laplace noise has been added, the skewed data is recorded on the blockchain through the smart contract. This perturbed data is saved by the web application temporarily. This is analogous to users sending data to the data curator by adding noise locally instead of trusting the data curator with their data. The blockchain can function as the data curator in this respect.
- Global Differential Privacy: The user sends a query through the frontend of the decentralized app, which is received by the web application as a POST request. The web application queries the original database, gets the result of the query, adds Laplace noise to it and thereafter returns the result to the decentralized app. The result is also stored on the web application temporarily. The decentralized app records the result through the smart contract on the Ethereum blockchain platform, which can be seen by the user who sent the original query as well as others who have a similar query. In this scenario the algorithm that computes the result of the query and adds noise to it functions as the curator.

V. Privacy Management Plane Implementation

The privacy management plane broadly consists of two components, namely the decentralized app, Laplace dapp, and the web application. An overview of the implementation is given in Fig. 2. The privacy management plane was developed using the Truffle development environment [22], Flask 1.1.1 [23] and Python 3.7.3. The smart contract was coded and tested on Remix [24], the local Ganache [22] blockchain network and public Ethereum testnet, Ropsten [25]. It is available for access at the Ropsten address *0x13D2E4931C821763145bf18e1f8bE87079F4c84C*. The cost computation was also accomplished on all the testing platforms and shown for Ropsten testnet in Table I. A video depicting the demo of privacy management plane on the local Ganache blockchain network can be accessed from [26].

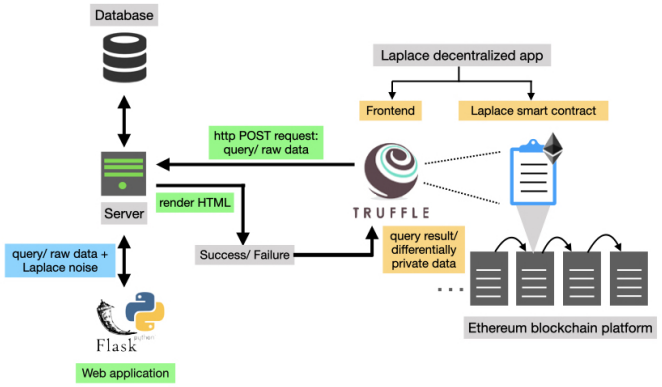

Fig. 2: Implementation Overview of the Privacy Management Plane

We used a dataset containing loan details linked to accounts, *loan.asc*, taken from an anonymized dataset of Czech bank [27] to implement and test the smart contract in the privacy management plane. The dataset had columns *loan_id, account_id, Date, Amount, Duration, Payments and Status*. The coded smart contract in the decentralized management plane has functions to cater to recording the query results for this dataset. The differentially private data input by the user to the smart contract to demonstrate local differential privacy incorporates utilizing the data from this dataset for the purpose of validation of the implementation. The data used is merely for the purpose of demonstration of the utility of the implementation in preserving privacy by skewing the original data. A function to answer a simple query by the user to find the *mean* of the column *Payments* or the column *Amount* is given in Listing 1.

Relevant code before the indicated function declares a new type through a *struct Query*, which includes 4 fields namely the ID of the query, the selection of the column the user desires the *mean* of, the value of ϵ and the query result. It must be noted however that in a practical deployment the value of ϵ should not be known to the user in the scenario of global differential privacy. We have included it in the smart contract to assist in the experimental evaluation of the developed dapp and to demonstrate by our development that blockchain can be utilized to achieve differential privacy. A *mapping* is also declared before the indicated function to ensure that for the ID of the query used as *key* we can access all the fields in the query. The most recent query will have an ID equal to the total number of queries and hence the usage of *numQueries* to access the fields in the query. Lines 7 to 15 comprise of the function to record the query result to find the mean of the desired column in the dataset.

```
1  pragma solidity 0.6.1;
2  contract Primary{
3  //code preceding the function
4  mapping(uint=>Query)queries;
5  uint numQueries;
6  //function for simple query
7  function simpleQuery(uint epsilon, uint choiceID,
       ↪ uint result)public{
8     numQueries++;
9     queries[numQueries].queryID=numQueries;
10    queries[numQueries].epsilon=epsilon;
11    queries[numQueries].selection=choiceID;
12    queries[numQueries].result=result;
13 }
14 //rest of the code
15 }
```

Listing 1: Smart Contract Function for a Simple Query

The addition of Laplace noise to the data of a user to achieve local differential privacy is depicted in Listing 2. The function receives the value of the variable through a POST request and thereafter uses *numpy* to add Laplace noise depending on the value of ϵ chosen by the user. The perturbed data is written to a file, which is retrieved by the Laplace dapp to record the data on the blockchain.

```
1  #code preceding the function definition
2  @app.route('/localdp',methods = ['POST'])
3  def localdp():
4    if request.method == 'POST':
5      a = int(request.form['amount'])
6      duration = int(request.form['duration'])
7      payments = int(request.form['payments'])
8      epsilon = float(request.form['epsilon'])
9      scale=1/ epsilon
10     s=np.random.laplace(0, scale, 1)
11     a=a+s
12     duration=duration+s
13     payments=payments+s
14     f=open("localdp.txt","w+")
15     f.write("%d %d %d %f\n" % (a,duration,
           ↪ payments,epsilon))
16     return "Success! Local Differential Privacy
           ↪ Achieved"
17 #rest of the code
```

Listing 2: Addition of Laplace Noise to User Data

VI. Experimental Evaluation of the Privacy Management Plane

We tested the privacy management plane with the dataset, (n=683) containing loan details linked to accounts, *loan.asc*, taken from the anonymized dataset of Czech bank [27]. We deployed and tested the smart contract in the Ropsten testnet of Ethereum. The cost computation for the deployment of the smart contract in Ropsten and the functions used in the smart contract are given in Table I. The cost for the deployment of the smart contract (SC) is \$0.11. We coded a function to calculate the *mean* of the column *amount* or the *mean* of the column *payments* giving the option to the user to choose the column. A query to find the sum of either column costs \$0.02. We coded a function in the smart contract to cater to a complex query where we find the column *amount* grouped by the columns *duration* and *payments*. The query result comprises of 3 columns with 683 rows. The cost of this query in the blockchain was \$18.59. The cost computation has been done using average confirmation time, where the mean time to confirm was approximately 1874 seconds and 125.8 blocks on the day of computation [28].

Account	Gas Used	Price (ETH)	Price ($)
SC deployment	570199	0.0006272	0.10662
Simple query	127092	0.0001398	0.02377
Complex query	145656*683	0.1094	18.59
LDP	148103	0.0001629	0.02769

TABLE I: Cost Computation In Ethereum

In order to aid in our demonstration of the experimental evaluation we used a small subset (n=10) of the entire dataset from Berka [27], which is given in Table II. The given set of records indicates a few selected columns from the original dataset.

We conducted the following tests through the decentralized app [26]:

- Query to find the mean of the column *Payments* at different values of ϵ and read the query result from the blockchain. The true mean of the column is 4571.2, which

Amount	Duration	Payments	Status
96396	12	8033	B
165960	36	4610	A
127080	60	2118	A
105804	36	2939	A
274740	60	4579	A
87840	24	3660	A
52788	12	4399	A
174744	24	7281	B
154416	48	3217	A
117024	24	4876	A

TABLE II: Demonstration Dataset

can be verified from the dataset in Table II. The test results are given in Table III.

- Query to find the mean of the column *Amount* at different values of ϵ and read the query result from the blockchain. The true mean of the column is 135679.2, which can be verified from the dataset in Table II. The test results are given in Table III.
- Complex query to find the column *Amount* grouped by the columns *Payments* and *Duration*. The results are depicted in Fig. 3.
- We simulate the input of values by 3 users by the input of the raw data from the first three rows of the dataset given in Table II. The data comprises of three variables from the columns *Payments*, *Amount* and *Duration* through the privacy management plane. The perturbed result is stored on the blockchain. The results are read from the blockchain and are visualized through the Laplace dapp in Fig. 4. The figure depicts the accomplishment of local differential privacy, where the results are stored on the blockchain and the raw data is not stored at any phase of the interaction between the different components of the privacy management plane.

ϵ	Mean	Column
0.01	4917.78	Payments
0.2	4574.29	Payments
0.5	4571.67	Payments
0.05	135701.82	Amount
0.3	135689.01	Amount
0.7	135679.11	Amount

TABLE III: Simple Query Execution in the Privacy Management Plane

QueryID	Row	Epsilon	Duration	Payments	Amount
2	1	0.2	10	4397	52786
2	2	0.2	10	8031	96394
2	3	0.2	22	3658	87838
2	4	0.2	22	4874	117022
2	5	0.2	22	7279	174742
2	6	0.2	34	2937	105802
2	7	0.2	34	4608	165958
2	8	0.2	46	3215	154414
2	9	0.2	58	2116	127078
2	10	0.2	58	4577	274738

Fig. 3: Complex Query result from the Blockchain

An evaluation of the developed privacy management plane depicts that both global and local differential privacy can be

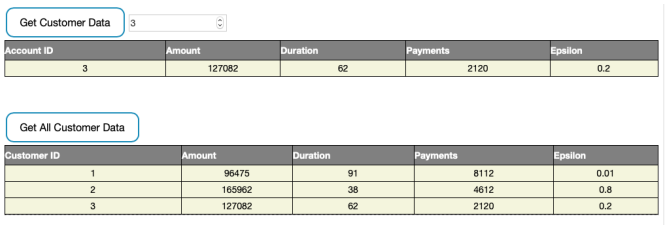

Fig. 4: Differentially Private Data read from the Blockchain

achieved through the blockchain ensuring the integrity of the recorded data. The complexity of the underlying blockchain and the privacy management plane will be hidden behind the decentralized application enhancing the potential usage.

A. Managing Smart Contract for Role-based Data Access

In [29], a management plane was developed, which facilitated data-filtering and monitoring services to blockchain-based applications employing smart contracts. The management plane accomplished role-based access to blockchain data through dedicated *managing* smart contracts. Their primary utility is that an organization can manage multiple smart contracts through a single managing smart contract. Blockchain permits data to be appended but prevents deletion ensuring that the data once added cannot be erased. We coded a managing smart contract (MSC) to manage the smart contract (SC) in the privacy management plane to restrict the read access to a blockchain record, through the SC. It will still be possible to read the data from the distributed database of the public blockchain. However in an optimally configured permissioned blockchain, the read access can be restricted to only through the smart contracts and then the target record will no longer be available to be read from the blockchain by the users. This can be accomplished by permitting only the users that register through the SC to have read access. A special identifier can be assigned to each registered user. Permissioned blockchains have an access-control layer built into the nodes and the implemented functionality in this paper to make a certain data record unavailable through the smart contract will work seamlessly [30].

```
1 pragma solidity ^0.6.1;
2 contract Managing{
3     function updateAccess(uint ID)public {
4       address SC=0xe7370Fd93bFF00e7Aa98c47665C7DD
            ↪ 18189CF5D2;
5       (bool success, bytes memory data) = SC.call(
            ↪ abi.encodePacked(bytes4(keccak256("
            ↪ updateID(uint256)")),ID));
6     }
7 }
```

Listing 3: Managing Smart Contract

The updated smart contract, SC, was deployed and tested on the Ropsten testnet of Ethereum. The address of the deployed smart contract is *0xe7370Fd93bFF00e7Aa98c47665C7DD18189CF5D2*. MSC was also deployed and tested on Ropsten and the address of the deployed smart contract is

0xE94442bAb9c6500f842E374D7013953ee240630c. We computed the costs of MSC and the updated smart contract, SC, which incorporated an additional function, *updateID(uint ID)*, to update the ID of the record accessed to revert the transaction in case the record with that ID is queried. MSC was provided with the address of the deployed smart contract, SC, and the function to update the target record was made *private* in SC. The price in $ for the transactions reflects the conversion rate the day the transaction was conducted and is subject to fluctuations. The mean time to confirm the transactions was 1874 seconds while the mean time to confirm in terms of blocks was 125.8 [28]. MSC is given

Account	Gas Used	Price (ETH)	Price ($)
Updated SC Deployment	605114	0.0006656	0.11315
MSC deployment	165493	0.000182	0.03094
updateID(uint ID)	22935	0.0000252	0.00428

TABLE IV: Cost Computation for Integration of a Managing Smart Contract

in Listing 3. Line 4 declares the Ropsten address of the deployed SC while line 5 calls the function *updateID(uint ID)* in the SC. As a result of this when a call is made to read the data of the customer with the ID passed from the managing smart contract, the blockchain reverts the transaction through the SC.

VII. Compliance to GDPR and Implications

The rise of digitising communication forms and societal relationships has positioned differential privacy as an important mechanism to tackle socioeconomic concerns [31]. Although this paper is targeting computational aspects of differential privacy, and introduces a novel approach to manage differential privacy in blockchains, the societal and economic implications of the developed privacy management plane are also an important consideration. Individuals' attitude on data privacy can vary over time and it is therefore of crucial importance to differentiate among different data privacy attitudes of individuals [32]. The Privacy Act in the United States [33], which dates back to 1974, follows a "Notice and Choice" policy to safeguard data privacy of individuals. The recently introduced General Data Protection Rule (GDPR) [34] in the European Union, however, extends upon this policy, and introduces the rule of "Right to be forgotten" to adopt regulatory mechanism to changing data privacy considerations of individuals, as well as to safeguard data privacy of individuals in the digitized world. GDPR establishes clear and consequent rules that govern the transfer and circulation of personal data among different parties, along data supply chains. It is a novel, forward-looking regulatory element of its kind. As such, it declares and provides elementary rules, assigning power to individuals to manage their own data according to their preferences. Consequently, the decision-making power that is being given implies new regulatory requirements that institutions, data curators need to seriously comply with and failure to adhere would incur high financial penalty [35].

Multiple points of tension have been identified between GDPR and blockchains and they can be broadly categorized into the following factors as causative agents [36]. The factors responsible for the discord between blockchains and GDPR together with an elaboration on how the privacy management plane resolves them are given below:

1) **GDPR is based on the assumption that for each data point, there exists a legal entity (data controller), who is responsible for the enforcement of the rights of the data subjects under the GDPR.** The developed privacy management plane while ensuring global differential privacy queries a database under the control of a data curator, who is responsible for determining the privacy budget and storing the perturbed query result on the blockchain. In case of local differential privacy, GDPR compliance can be achieved in a permissioned blockchain network, where the legal entity/ entities who employed a permissioned blockchain network for their organization are responsible for the enforcement of GDPR. Their role can be seen in the managing smart contract that was coded to prevent a data record on the blockchain from read access.
2) **GDPR is based on the assumption that data can be erased or modified when necessary in compliance with the legal requirements such as Articles 16 and 17 GDPR.** The data queried from the database in global differential privacy is off the blockchain and any record can be easily erased or modified. The query result stored on the blockchain will not be impacted significantly by the addition or deletion of a data record pertaining to an individual in the database as per the definition of differential privacy discussed in section III. In case of local differential privacy, we state the significance of the access layer in a permissioned blockchain network for the implementation of the privacy management plane discussed in subsection VI-A. The managing smart contract when given the ID of the customer record stored on the blockchain updates the read access to *revert()* the transaction when the ID is invoked using a *call()* preventing the record from being read.

Additionally in achieving global differential privacy the data curator has already anonymized the dataset before Laplace noise is added to the query result through the privacy management plane. In case of local differential privacy, the pseudonymous identity of the individuals revealed through the public-private key pair can be masked by a single or multiple designated blockchain addresses recording data on behalf of the individuals in the blockchain. The cost computation of the transactions conducted by an individual can be negotiated off the blockchain with deduction of deposited fiat money/ cryptocurrency on registration in the decentralized app and input of data through the decentralized app before it is perturbed to be stored on the blockchain.

The significance of the developed privacy management plane in adhering to GDPR lies in a crucial design element,

where the employment of the privacy management plane to store data on blockchains becomes the function of the differential privacy configurations of the data owner himself. This architectural design strategy provides competitive advantage for any blockchain-based application especially for the European market, as it ensures compliance to GDPR, permitting data to be shared only if the individual, i.e. the data owner, allows it. As a novel socioeconomic consequence, the proposed solution actually gives the decision-making power directly in the hands of users to manage the privacy of their own data. Besides the associated operational costs of utilizing the privacy management plane, the differential privacy choices of data owners might carry potential financial implications for the data curators too. This added power thus might allow data curators to introduce novel data handling practices for users with direct commercial and socioeconomic consequences. Our proposed application for safeguarding digital information upon individuals' choices will allow corporations to develop blockchain-based applications that not only remain compliant with elementary data protection rules, but also pave the way for developing and applying incentive schemes for individuals. This will facilitate individuals to address and to handle their data as an asset. On a longer run, such a mechanism could lead individuals to explore and to understand the valuing fundamentals of their own personal data.

VIII. Conclusion

In this paper, we introduced a novel privacy management plane which is a pioneering step towards achieving GDPR compliance through the blockchain using differential privacy managed through smart contracts. The developed privacy management plane facilitated data curators to allow queries on a database through a decentralized application with the query results being stored on the blockchain. The storage of query results on the blockchain paves the way for multiple usage of the results without compromising on the integrity of the result and absolves the need for repetition of the same query by others. The privacy management plane also demonstrated the integration of local differential privacy with the individual data owners selecting the level of data perturbation and controlling the privacy budget through a blockchain-based decentralized application. Cost computation was done to give an estimate of the financial expenses that will be incurred to store data on the blockchain using differential privacy through a smart contract. An extension to the privacy management plane integrated the management by another smart contract to prevent a data record stored on the blockchain from being read through the smart contract in the privacy management plane. This enhancement when used in a permissioned blockchain network will ensure the data owner's right to revoke his permission for sharing his data. The developed decentralized application to achieve local differential privacy can be used by organizations to collect perturbed data, giving the users the right to control the privacy of their data with the data being stored off the blockchain if needed and queried in a global privacy context through the blockchain. The implementation is a basic step towards ensuring data privacy and compliance to GDPR. Future work will involve developing a more secure mechanism to access the relational database and add Laplace noise.

References

[1] S. Chen, J. Zhang, R. Shi, J. Yan, and Q. ke. (2018, 01) A comparative testing on performance of blockchain and relational database: Foundation for applying smart technology into current business systems.

[2] N. Lindsey, "Blockchain Privacy Poisoning a New Concern in Post-GDPR Era," https://www.cpomagazine.com/data-protection/blockchain-privacy-poisoning-a-new-concern-in-post-gdpr-era/, accessed: 6th April, 2020.

[3] H. Kenyon, "Privacy 'Poisoning' Cyberattacks Pose Risk to Blockchain," https://www.govtech.com/security/Privacy-Poisoning-Cyberattacks-Pose-Risk-to-Blockchain.html, accessed: 6th April, 2020.

[4] H. Kalodner, S. Goldfeder, A. Chator, M. Möser, and A. Narayanan. (2017, 09) Blocksci: Design and applications of a blockchain analysis platform.

[5] M. Möser, K. Soska, E. Heilman, K. Lee, H. Heffan, S. Srivastava, K. Hogan, J. Hennessey, A. Miller, A. Narayanan, and N. Christin, "An empirical analysis of traceability in the monero blockchain," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, pp. 143–163, 06 2018.

[6] R. Hackett. (2019) Zcash discloses vulnerability that could have allowed 'infinite counterfeit' cryptocurrency. [Online]. Available: http://fortune.com/2019/02/05/zcash-vulnerability-cryptocurrency/

[7] G. Zyskind, O. Nathan, and A. . Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops*, 2015, pp. 180–184.

[8] X. Chen, J. Ji, C. Luo, W. Liao, and P. Li, "When machine learning meets blockchain: A decentralized, privacy-preserving and secure design," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 1178–1187.

[9] M. Ul Hassan, M. H. Rehmani, and J. Chen, "Privacy preservation in blockchain based iot systems: Integration issues, prospects, challenges, and future research directions," *Future Generation Computer Systems*, vol. 97, 03 2019.

[10] M. Yang, A. Margheri, R. Hu, and V. Sassone, "Differentially private data sharing in a cloud federation with blockchain," *IEEE Cloud Computing*, vol. 5, no. 6, pp. 69–79, 2018.

[11] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustainable cities and society*, vol. 39, pp. 283–297, 2018.

[12] N. Khan and M. Nassar, "A look into privacy-preserving blockchains," in *16th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2019*, 08 2019.

[13] I. Dinur and K. Nissim, "Revealing information while preserving privacy," 01 2003, pp. 202–210.

[14] C. Dwork and K. Nissim, "Privacy-preserving datamining on vertically partitioned databases," in *Advances in Cryptology – CRYPTO 2004*, M. Franklin, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 528–544.

[15] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: The sulq framework," 01 2005, pp. 128–138.

[16] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5018–5029, 2016.

[17] S. Vadhan, *The Complexity of Differential Privacy*, 04 2017, pp. 347–450.

[18] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang, "Privacy at scale: Local differential privacy in practice," in *Proceedings of the 2018 International Conference on Management of Data*, ser. SIGMOD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1655–1658. [Online]. Available: https://doi.org/10.1145/3183713.3197390

[19] X. Hu, M. Yuan, J. Yao, Y. Deng, L. Chen, Q. Yang, H. Guan, and J. Zeng, "Differential privacy in telco big data platform," *Proc. VLDB Endow.*, vol. 8, no. 12, p. 1692–1703, Aug. 2015. [Online]. Available: https://doi.org/10.14778/2824032.2824067

[20] P. Jain, M. Gyanchandani, and N. Khare, "Differential privacy: its technological prescriptive using big data," *Journal of Big Data*, vol. 5, 12 2018.

[21] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, p. 211–407, Aug. 2014. [Online]. Available: https://doi.org/10.1561/0400000042
[22] "Truffle Suite," https://www.trufflesuite.com, accessed: 31st March, 2020.
[23] "Flask," https://flask.palletsprojects.com/en/1.1.x/, accessed: 31st March, 2020.
[24] "Remix - Ethereum IDE," https://remix.ethereum.org, accessed: 31st March, 2020.
[25] "Etherscan: Ropsten Testnet Explorer," https://ropsten.etherscan.io, accessed: 7th April, 2020.
[26] "Differential Privacy Enforcement through Ethereum," https://www.youtube.com/watch?v=RtXJHlGqM2U, accessed: 31st March, 2020.
[27] "The Berka Dataset Visualization," https://ensquad.com/2018/06/21/the-berka-dataset-visualisation/, accessed: 15th March, 2020.
[28] "ETH Gas Station," https://ethgasstation.info/calculatorTxV.php, accessed: 7th April, 2020.
[29] N. Khan, A. Lahmadi, J. Francois, and R. State, "Towards a management plane for smart contracts: Ethereum case study," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1–6.
[30] P. Novotny, Q. Zhang, R. Hull, S. Baset, J. Laredo, R. Vaculín, D. Ford, and D. Dillenberger, "Permissioned blockchain technologies for academic publishing," *Information Services Use*, vol. 38, pp. 1–13, 01 2018.
[31] "Researchers finally get access to data on Facebook's role in political discourse," https://www.sciencemag.org/news/2020/02/researchers-finally-get-access-data-facebook-s-role-political-discourse, posted in Social Sciences, 13 February, 2020.
[32] K. Nissim, T. Steinke, A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, D. R. O'Brien, and S. Vadhan, "Differential privacy: A primer for a non-technical audience," in *Privacy Law Scholars Conf*, 2017.
[33] M. Hulett, "Privacy and the freedom of Information Act," *Administrative Law Review*, pp. 275–294, 1975.
[34] "Complete guide to GDPR compliance," https://gdpr.eu, accessed: 3rd April, 2020.
[35] "What are the GDPR fines?" https://gdpr.eu/fines/, accessed: 3rd April, 2020.
[36] EPRS | European Parliamentary Research Service, "Blockchain and the general data protection regulation. Can distributed ledgers be squared with european data protection law?" *Scientific Foresight Unit (STOA), PE 634.445*, 2019.